Navy Personnel Research, Studies, and Technology Division
Bureau of Naval Personnel (NPRST/PERS-1)

Millington, TN 38055-1000 NPRST-TR-06-1 September 2006

Civilian Turnover and Retirement Projection Study

Rodney S. Myers, M.S. · Tanja F. Blackstone, Ph.D.
Navy Personnel Research, Studies, and Technology.

Myron C. Murset, M.S.
iBaset

NPRST
research at work

NPRST-TR-06-1
September 2006

Civilian Turnover and Retirement Projection Study

Rodney S. Myers, M.S.
Tanja F. Blackstone, Ph.D.
Navy Personnel Research, Studies, and Technology

Myron C. Murset, M.S.
iBase t

Reviewed and Approved by
David M. Cashbaugh
Institute for Personnel Planning and Policy Analysis

Released by
David L. Alderton, Ph.D.
Director

Navy Personnel Research, Studies, and Technology
Bureau of Naval Personnel
5720 Integrity Drive
Millington, Tennessee 38055-1000
www.nprst.navy.mil

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
30-09-2006

2. REPORT TYPE
Technical Report

3. DATES COVERED *(From - To)*
10-1-03 to 9-30-04

4. TITLE AND SUBTITLE
Civilian Turnover and Retirement Projection Model Study

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Rodney S. Myers, Myron C. Murset, Tanja F. Blackstone

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Navy Personnel Research, Studies, and Technology (NPRST/PERS-1)
Bureau of Naval Personnel
5720 Integrity Dr.
Millington, TN 38055-1000

8. PERFORMING ORGANIZATION REPORT NUMBER
NPRST-TR-06-1

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Bureau of Naval Personnel
Civilian Community Management Division
1301 South Gate Road
Arlington, VA 20370

10. SPONSOR/MONITOR'S ACRONYM(S)
BUPERS, N11

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

A - Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The Civilian Community Management Division's (N11) most critical function is forecasting future workforce needs. This function requires predicting the state of the future workforce to provide precise and accurate metrics on civilian workforce needs (i.e., recruitment, appropriate size of developmental programs, workforce shaping, etc.). N11 has the opportunity to benefit from proven business processes from the active duty Navy, as well as develop business processes that could prove beneficial to the active duty Navy.

15. SUBJECT TERMS

Civilian, forecasting, retirement, resignation, industrial, science and engineering, administration, NAPPE, CIVFORS, time series, regression, workforce, Navy

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON / 19b. TELEPHONE NUMBER *(Include area code)*
UNCLASS	UNCLASS	UNCLASS	UNCLASS	35	Genni Arledge / 901-874-2115 (882)

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Foreword

This report was prepared as part of the Civilian Turnover and Retirement Projection Model Verification, Validation, and Accreditation project (Program Element 0605152N) sponsored by the Civilian Community Management Division (N11). The work described here was performed during Fiscal Year (FY) 2004, by the Navy Personnel Research, Studies, and Technology Division's Institute for Personnel Planning and Policy Analysis (PERS-11).

The objective was to perform a thorough evaluation of the Navy Pay Predictor Enlisted and the Army's Civilian Forecasting System models to determine whether the approaches used in these models are useful for Navy civilian community management and planning. A secondary objective was to develop initial specifications for a prototype model to enhance civilian community managers' ability to manage and plan the Navy civilian workforce.

We especially thank Mr. Josh Fowler (N11) for constructive input during this study, and wish N11 success as they take on the responsibility of managing Navy civilian personnel. Special thanks to Mrs. Geetha Mandava and Mr. James Woods (University of Memphis interns) for their contributions.

DAVID L. ALDERTON, Ph. D.
Director

Executive Summary

Accurate workforce analysis is especially critical given heightened security threats, evolving missions, rapidly evolving technology, and changing personnel demographics. Navy Personnel Research, Studies, and Technology (NPRST) appreciates the opportunity to study and provide recommendations to the Civilian Community Management division. The findings and recommendations discussed in this report were derived from months of technology and literature reviews, empirical data analysis, discussions with analysts, and in-house subject matter knowledge and experience.

Although new organizations are faced with multiple challenges, each capable of limiting its forward progress, there are advantages that new-starts have, such as, the ability to perform analysis and strategic planning prior to implementation and operation. Once in operation, business processes become difficult to change (re-engineer) simply because they have been functional for a period of time. Modeling and simulation is especially useful to gain insight for planning and strategy. Our hope is for the Civilian Community Management division to utilize these results for future planning and strategic analysis.

The Navy's Pay Predictor Model for Enlisted (NAPPE) was evaluated for potential utility for civilian workforce planning. Although the current NAPPE model cannot support civilian workforce planning, given the differences between the military and civilian data, there are parts of the model that are useful. These are discussed in detail.

The Army's civilian forecasting system (CIVFORS) was evaluated for potential utility for civilian workforce planning. There are several parts of this model that apply to the functional requirements of a Navy civilian workforce-planning tool. These are discussed later in this report.

This study did not uncover any new forecasting techniques, but rather confirmed existing techniques. The challenge is identifying the most theoretically correct technique. We recommend consideration of the following when choosing a forecasting method: variables being forecast, availability of historical data, length of the forecast time horizon, and the aptitude of the analysts. We concluded that accuracy measures are unique to individual organizations and have little or no credibility beyond the organization. The baseline measure of forecast accuracy should be forecasts derived from a naïve method, and any other method should pass/fail based on how its accuracy compares to forecasts generated using the naïve method.

An extensive analysis of three forecasting methods was conducted; naïve, time-series, and regression. The results of each are discussed later.

Recommend the Civilian Community Management division obtain a decision support tool(s) to enhance its community managers' analysis, problem solving, and decision-making abilities. This tool should be built as a prototype prior to full development—this will allow civilian community managers to contribute to its development, and orient themselves in phases. Existing methods and business processes should be understood and utilized from other organizations where feasible. Prior to receiving such a tool, recommend civilian community managers be selected based on

their experience with analytic techniques. We recommend the use of Microsoft Excel as an interim environment for analysis and reporting because of its approval and support within the Navy/Marine Corps Internet (NMCI) environment and its analytical functionality. Any application developed by community managers during this interim phase should be shared with the prototype development team for incorporation in the prototype decision support tool.

Specifications for a prototype model are discussed in detail later in this report. These specifications should serve as a flexible baseline for the development of a decision support tool(s) for Navy civilian workforce planning.

Contents

List of Tables

Introduction

Abstract

The Civilian Community Management Division's (N11) most critical function is forecasting future workforce needs. This function requires predicting the state of the future workforce to provide precise and accurate metrics on civilian workforce needs (i.e., recruitment, appropriate size of developmental programs, workforce shaping, etc.).

N11 has the opportunity to benefit from proven business processes from the active duty Navy, as well as develop business processes that could prove beneficial to the active duty Navy.

Objective

The primary objective was to perform a thorough evaluation of the NAPPE and CIVFORS models to determine whether their prediction methods are suitable for predicting changes in the Navy's civilian workforce. A secondary objective was to develop specifications for an initial prototype model.

Background

The roles and responsibilities of N11 include

- Attracting, developing, and sustaining a diverse civilian workforce to meet the Department of the Navy's (DON) evolving mission requirements.
- Fostering a sense of civilian community across DON.
- Providing individuals the opportunity to develop to their fullest potential.

N11 began functioning at the beginning of fiscal year 2003. For additional information pertaining to the Civilian Community Management Division, visit http://www.donhr.navy.mil/ccm/.

Model Analysis

Navy Pay Predictor—Enlisted (NAPPE)

The NAPPE system contains two components, the user interface written in Visual Basic and the computation engine written in FORTRAN. The files used by the system are separated into three subdirectories: DATA, OUTPUT, and SCENARIO.

There are four different files of force structures: U.S. Navy (USN), U.S. Navy Reserve (UNSR), All Navy (ALNAV), and All Navy Totals (ALNAVTOT). The first three files contain 32 records for each quarter. These records represent the 31 Length of Service categories and a total. The ALNAVTOT file contains only the total lines from ALNAV.

The four files are updated quarterly with each update appended to the existing data. It calculates the values for basic pay, mean Length of Service (LOS), and continuation rates for the new quarter and appends these values to their corresponding history file. Each file has a header record indicating the fields contained in the file as a reference when viewed. This record is skipped by the FORTRAN module.

There are two record types in each pay table in the history file. The first is a data header record followed by 32 detail records. The header record contains the month, day, and year that this table becomes effective. The detail records contain monthly basic pay (in pennies) for nine pay grades and one record for each LOS cell.

NAPPE produces a summary report and two standard reports, Projected Basic Pay and Projected Mean LOS; along with an optional report, Projected Force Structure (for each quarter requested). These force matrices can be output as a text file for use as input to other programs. The Visual Basic front-end builds and displays reports using the existing reports and other history and output files.

The analyst may request special validation runs that produce various statistics used to compare system predictions with actual data. A validation run produces four summary reports of its own (Summary of Error Percentages, Summary of Mean LOS Differences, Average Errors in Costs, and Average Error in Mean LOS) plus the three reports mentioned above.

Managing the Navy's manpower requirements is a very complex balance between satisfying job requirements while meeting the fiscal constraints of an annual budget. The Navy's financial support division (N10) uses NAPPE to determine the feasibility of future manning requirements by estimating the cost of basic pay. It uses historical and current data to help budget analysts track budget execution in the current year, allowing managers to make any adjustments necessary to stay within the budget, or to request additional appropriations if needed.

Basic pay for individuals is determined by their military rank (referred to as pay grade), and their LOS. A table representing a count of all enlisted personnel in the Navy is called the force structure. The NAPPE forecasting system is used to forecast the enlisted force structure and the cost of basic pay. A more complete description of the problem was prepared by the Navy Personnel Research and Development Center in 1977 (Chipman, 1977)[1].

Background

The force structure is a standard matrix containing the counts of personnel by pay grade and LOS in years. The margins are the total number of members in these dimensions. There are 9 pay grades, labeled as E-1 to E-9, plus the total; and 31 LOS bins, one for each for each year of service, a cumulative category for 30 years or more plus the total.

[1] Chipman, M. (1977). *Forecasting the Naval Enlisted Personnel Force Structure to Estimate Basic Pay* (NPRDC-TR-78-4). San Diego: Navy Personnel Research and Development Center.

Pay grades E-1 to E-3 are the un-rated force, grades E-4 to E-6 are Petty Officers, and grades E-7 to E-9 are Chief Petty Officers. LOS cells 1 to 4 contain personnel in their first term of enlistment. Individuals with five or more years of service are referred to as careerists. The official pay tables for enlisted basic pay are the same for all services (i.e., Army, Navy, Air Force, Marines, etc.).

NAPPE uses historical force structures, current year force structure, and future strength plans to develop a forecast. Calculating basic pay is a simple process of multiplying the appropriate pay table rate by each of the counts (pay grade by length of service) in the force structure. NAPPE does not currently forecast additional pay categories, such as: basic allowance for housing (BAH), Assignment Incentive Pay (AIP), Pro Pay, etc. Choosing the method of forecasting is not as simple.

The nature of the underlying processes to determine how any pay grade or length of service count changes over time is extremely complex. Manpower requirements are driven by positions that must be filled. There are regulations regarding length of tours at sea and ashore, and rules regarding promotion. There are ebbs and flows in the pay rate compared to the private sector that affect each choice an individual Sailor makes during his/her military career.

There are generally two ways to approach this type of forecasting problem. One is to model the flow of people into and out of each cell in the force structure. This has been used in the past with some success. It requires accounting for future military requirements, which can be driven by events (social, political, and economic), that are difficult to predict. The number of ways people can enter and leave a LOS cell complicates a flow model for this process. The developers of the NAPPE model considered 20 years of historical data and found that it was highly seasonal; the numbers fluctuate depending on the time of year.

They realized that using a flow model would be deterministic, and coupled with incomplete knowledge of external variables, would likely yield unsatisfactory results. They decided on a two-pronged approach using only historical data to forecast the length of service margin of the force structure. Given the pay grade margin from the desired future strength, it is possible to find the individual cell counts inside the matrix that fit the margins. The details of how this is accomplished, including its verification and validation follows.

Mathematical Model

Forecasting the Length of Service Margin

Within NAPPE, three sets of data are available and maintained going back to 1957 for Navy enlisted personnel. These represent three force structures: U.S. Navy (USN), U.S. Navy Reserve (USNR), and All Navy (ALNAV). The USN matrices contain the counts of regular and active duty personnel. USNR is the count of reservists on active duty. ALNAV is the sum of USN and USNR. These files are updated quarterly. Other historical data used are the pay tables for all the years in the history files.

In determining the best approaches to forecast manning levels, developers discovered two possible ways to characterize the changes over time to each LOS cell: the change in raw counts, or a rate of change. This rate of change is defined for modeling purposes as the continuation rate (i.e., the proportion of people in a given LOS that move to the next LOS cell in the next time interval). After experimentation, it was concluded that in some cases the forecast using the raw count variable yielded the best forecast; in others, the rate variable was superior. The decision was made to use both variables in modeling the data. The model runs two parallel sets of forecasts for each of the three input data sets and chooses the forecast for each LOS cell with the least error. For a complete discussion of this analysis see Chipman (1977).[2]

The model uses time series techniques to forecast the LOS margin of three populations: USN, USNR, and ALNAV. It uses various forms of the Autoregressive Integrated Moving Average (ARIMA) process (Box & Jenkins, 1970; Brown, 1963). The model set uses autocorrelation, exponential smoothing, seasonality, and moving averages, in all combinations and with varying coefficients for each of the 90 LOS cells and the 2 variables. The analysis of the data prior to developing the actual model showed ARIMA is likely to yield the best results for each of the 90 LOS cells using data from 1957 to 1977. Tables in Chipman (1977)[3] tabulate these results for each of the two variables.

Cells 1 and 31, as is often the case with end-points, require special handling. Cell 31 is the only LOS cell containing multiple years. Some in Cell 31 will remain through the next time interval, while others will arrive from LOS 30. A special rate of change for Cell 31 is defined to include the personnel that remain. See equation 7 in Chipman (1977).[4]

Cell 1 is also a special case because there is no preceding cell from which to compute the two variables. This cell is also volatile due to the varying numbers of recruits coming in at any point in time. Since the total end-strength is given as input from the strength plan, Cell 1 must equal the end-strength minus the sum of counts in the other 30 cells. This is referred to as a residual forecast. There may be some special cases where the number of recruits is abnormally high or low, and thus the residual forecast may not be realistic. Research of the statistical properties of Cell 1 in the historical data yielded a set of boundary constraints that can dampen the sudden change in recruits. The model computes the difference between the residual estimate and a dampened estimate for Cell 1. If there is a difference, the discrepancy is redistributed to other LOS cells in such a way as to minimize the predictive error.

Building the Interior of the Force Structure

The interior of the (standard) force matrix is obtained by using the given pay grade margin (obtained from strength plans), the forecasted LOS for these individuals at this time, and the distribution of a standard matrix. This is an iterative method for finding the entries of a standard matrix given the margins and the characteristics (proportions) of a reference matrix. It alternately adjusts the interior of two matrices working across

[2] Ibid., p. 5.
[3] Ibid., p. 7–9.
[4] Ibid., p. 11.

and down, then re-normalizes until the two marginal totals match. The obvious choice for a reference matrix would be the most recent actual force structure. Again, careful analysis of the historical data concluded a better result could be obtained by using a composite matrix computed from the last 3 years of data. Chipman (1977)[5] includes a complete discussion of this topic as well as the equation for building the reference matrix.

Verification and Validation

Verification and validation procedures are incorporated into the model. The verification procedures provided feedback during program testing for debugging and insured the modules were operating correctly. The validation is included in the program, because the original version was an experiential prototype. The measurements of errors of estimation are a part of the internal logic of the model. Validation is a special case of running historical data though the forecast then adding special reports that compare the forecasts with the actual data.

The original validation results are reported in Chipman (1977).[6] The overall mean error was 0.24 percent of cost for a 1-year forecast, 0.76 percent for 2 years and 1.03 percent for 3 years. The validation includes turbulent years from the Vietnam conflict. The model performs better in more stable times.

Civilian Forecast System (CIVFORS)

The Army's civilian forecasting system has undergone two comprehensive verification, validation, and accreditations (VV&A), as well as evaluation by the Government's Accounting Office (GAO). This section discusses CIVFORS relative to its intended purpose and offers a recommendation about its applicability to Navy civilian workforce planning. Much of the underlying background and explanation related to CIVFORS was taken from the U.S. Army.[7]

The Army uses CIVFORS to plan its civilian workforce. CIVFORS supports several planning initiatives: annual evaluations, studies of workforce dynamics in hiring and separations, retirement projections, workforce realignment assessments, projected intern requirements, evaluation of separation incentives, studies of turnover by occupation, hiring plans, etc. These functions reside under the Army's workforce planning function. The use of CIVFORS in this context is to help decision makers anticipate what is likely to happen, under certain conditions. The Army is currently using CIVFORS for systematically gathering all projected hiring requirements across the Army to institute an Army-wide perspective that can be used to yield a strategic approach to filling needs.

[5] Ibid., p. 13–14.
[6] Ibid., p. 17.
[7] U.S. Army. (2003). *Civilian Forecasting System Verification and Validation Report.*

CIVFORS has two audiences: viewers and builders. Viewers of published forecasts use the information as needed and represent the largest portion. The majority of viewers are personnel specialists. The builder audience consists of analysts who develop models and make them available for viewing.

CIVFORS was originally developed and accredited in 1987. The original system was hosted on an IBM mainframe computer, but has since been rewritten in more efficient computer programming languages. The mathematical modeling and forecasting methods defined in the original CIVFORS design were not affected. Enhancements have been made in terms of program consolidation, object oriented design, database management, program flexibility, and graphical user interface. Today, CIVFORS is a web-based system that connects the Army Civilian Personnel Management System (CPMS) through the Internet. Hardware for the CIVFORS consists of a database server, an application server, and a web-server, all of which are Intel-based Pentium processors running Windows NT.

It consists of six modules (Rate Generator, History Generator, Inventory Projection, Linear Optimization, Target Generator, Data Mining). The heart of the projection accuracies rests with the quality of the methodologies used in the Rate Generator and Inventory Projection modules.

Per the *Civilian Forecasting System Verification and Validation Report* (October 2003), the overall findings unequivocally show CIVFORS performance in all required testing areas continues to be well established. They report the Army has taken the steps necessary to ensure the model's forecasting accuracy for the Army's civilian workforce.

CIVFORS Capabilities

Comparisons of projected to actual data over sequential years show the aggregate and detail accuracy levels are well within the expected ranges of 90–100 percent. The results included all projected to actual comparisons despite any abrupt, unusual, and sudden changes that were unknown to the model for the forecasted period such as realignments, reorganizations, abruptly introduced separation incentives, major policy changes, etc. Including the unknown factors in testing has both drawbacks and advantages. The drawback is that no model can perfectly predict the future inclusive of anomalies. Hence, it is an unfair assessment to expect the model to incorporate anomalies it does not know that may be forthcoming. The adoption of this testing approach is realistic because in the "real" world the unexpected happens and the model performance in the "real" world is ultimately what matters.

CIVFORS is a fully web-enabled automated analytic tool used in "flexible" model construction. It mirrors existing Army enlisted, officer, and civilian systems. The forecasts are based on past or future assumptions with or without optimization at the Army, Major Command (MACOM), installation, and Unit Identification Code (UIC) level of detail, provided necessary data are available. Data sources within CIVFORS are managed in an ORACLE database system containing more than 4 million transaction records. For each run of the production models, the system generates more than 200 megabytes of numerical data to support the forecasting and viewing components.

CIVFORS reports along 12 different dimensions containing more than 700,000 strength records in history and projections with all combinations available for query and review. Projections (7-year forecast) are based on the previous five years of historical data.

CIVFORS is able to perform the following types of projections:

Baseline—No Goal: if the past 5 years of history were to repeat itself where would we end up in the next 7 years?

Baseline—Goal: if given a target projection based on budgeted and authorized spaces and if history were to repeat itself, can we meet the target projection for the next 7 years?

What If—Simple Assumption: if future rates were different (i.e., history did not repeat itself against a given set of targets) what would happen over the next 7 years?

What If—Complex/Multiple Assumptions: "what ifs" can be translated into specific terms that the model can handle, e.g.:

1. Salaries doubled in some series
2. Contracted out
3. Automated more
4. Changed the mix within budgeted/authorized limits

As an overview, the historical data are built in "time series" format along dimensions determined by data mining analyses or comparative analyses to be the most predictive for gain, loss, and migration transactions included in the model. The historical rates are computed along the model's predictor dimensions by type of transaction for gains, losses, and internal movements with aging rates factored in the computations (e.g., voluntary loss rates by agency, Professional, Administrative, Technical, Clerical, and Other or Blue Collar jobs) by years of service, employee tenure, and retirement eligibility groups. The rates are forecasted into the future by extrapolating the historical rates using time series. For the no goal and goal forecasts the rates are computed and applied to update inventory for each time period of the projection in the following order: (a) movement into and out of age groups, (b) gains from outside the workforce (e.g., new hires), (c) movement within the workforce (e.g., agency transfers, occupation changes, promotions), and (d) separations from the workforce (e.g., retirements, voluntary separations). The no goal forecasts are based strictly on the historical patterns and trends extrapolated into the future. The goal forecasts include an optimization step that computes gains and losses to the workforce based on future needs. For the no-goal and goal forecasts, the future size of the workforce is computed in the order above. The workforce is divided into groups based on similar demographic job categories (e.g., agency, occupation, gender, etc.) so that appropriate transaction rates can be applied.

CIVFORS Modules

The rate generator produces projected transaction rates that are used by the inventory projection module to determine future transaction counts and strength levels. It then examines the historical rates to detect outliers that will not be used in the

calculations to determine projected rates. The rate generator uses one of five rate extrapolation routines to determine projected transaction rates for each cell. The rate generator determines historical rates based on strength and transaction count history for each cell based on one of two methods. The weighted average and seasonal weighted average routines are determined based on simple division of transaction count divided by strength. For other routines, it uses a hierarchical rate-blending algorithm.

For a given cell, the algorithm begins with a 1-dimensional rate based on the most significant dimension (as determined by the rate generator). Using a weighting scheme this rate is blended with the 2-dimensional rate based on the next most significant dimension. This process continues until there are no more dimensions that significantly alter the rate. This methodology is quite effective in establishing rates for populations with small, medium, and large cell sizes. For large cells, no rate blending occurs since there are sufficient data upon which to base a rate. For medium cells, rate blending occurs according to the algorithm described above. For small cells (typically with strength less than 25), rate blending occurs with much greater weight on the higher-level rates calculated. As a result, small cells often have transaction rates similar to other small cells having many of the same dimensional values.

The inventory projection module uses a life-cycle modeling process to compute projected strength and transaction counts based on rates produced by the rate generator and any optimized transaction data produced by the optimization component. CIVFORS models workforces in a given time period by following these sequential steps: (1) age the workforce by considering migration transactions based on changes in time-based data elements such as years of service or age, (2) add all gains, (3) subtract/add those moving from one cell to another within the population due to a migration transaction based on a change in non-time based data element, and (4) subtract all losses. This process is carried through the computations in both the inventory projection module and the AMPL (a modeling and scripting language) optimization model.

For a goal projection, targets are needed for the mathematical goal optimization inventory projection modules. The targets must be provided at the level of detail identified in each module. The target generator module processes data from systems such as the Army's budgetary and force planning system. Once processed, these data are distributed into aggregate counts by quarter for each fiscal year of the program objective memorandum (POM) and across various modeled dimension combinations (i.e., occupational series, pay plans, grades, etc.). These quarterly targets are required for both the goal optimization and the inventory projection module. Quarterly targets are not interpolated in the target generator. Targets for each projection period are generated through a direct assignment from the target database.

Potential Uses

The NAPPE data is generally non-stationary. There is seasonality in many of the cells, especially the USN and ALNAV data. USNR personnel behave differently, so are modeled separately to improve the total force forecast.

Relative to the Navy's civilian workforce NAPPE only addresses two populations (USN and USNR), one pay system, and a single level of aggregation. The civilian force is composed of several different pay systems; multiple retirement systems; and industrial, professional, and executive populations.

NAPPE is a dynamic model that has survived major changes to the Navy's workforce. It proves statistical forecasts of a large, relatively slow to changing population can be very accurate and useful. The civilian force is more complex than the Navy's enlisted force. One might apply some of NAPPE's methods to subsets of the civilian force, but care must be taken to ensure there are sufficient sample points and counts in each category to provide accurate results that meet the minimum statistical requirements of an accurate time series forecast.

Analysis of CIVFORS and its verification and validation processes did not reveal any non-trustworthy practices. The comprehensive aspect of how the model represents the "real" world and the extent it has undergone objective verification and validation was impressive. However, the current CIVFORS will not support Navy civilian workforce planning for several reasons: (1) data and variables do not match across Army and Navy, (2) a data mining effort must be utilized for handling input data, (3) hardware and software conflicts need to be resolved—particularly as it relates to NMCI. We recommend CIVFORS as a development guide for a Navy civilian workforce planning tool. We expect similar modules will prove necessary for a Navy model, however, this must be verified by the model development team. The various methods for generating rates will prove useful for the group of civilian community managers.

Industry Forecasting Techniques and Standards

Accurate forecasting is key in any strategic or tactical decision to ensure an organization's competitiveness. Although not all organizations have an active forecasting group, many organizations invest in estimating their future. As part of this study, other organizations were investigated to determine their approaches to predicting the future. Our search involved a review of technical forecasting literature via libraries, Internet searches via www.google.com, conversations with practicing forecasters, and attendance at professional conferences (e.g., Supply Chain Forecasting).

The goal was to identify mathematical techniques of generating forecasts, standard measures or thresholds for determining forecast accuracy, organizational best practices, and relevant commercial forecasting software packages.

Other such forecasting methods identified were conformation of common forecasting methods such as: regression, time series, extrapolation, etc. A glossary of these methods is documented in Appendix A. Per the sponsor's request: we are providing a more detailed evaluation of regression analysis.

The conclusions on forecast accuracy were (1) there is NO industry standard, (2) beware of using accuracy measures from other organizations without totally understanding their metrics (e.g., levels of aggregation, time periods, etc.), and (3) the challenge is to produce a forecast that is more accurate than one produced using the "Naïve" method. Again, this is a challenge—the ultimate recommendation is to use the most theoretically correct technique that produces the best forecast accuracy for organizational objective(s).

In reality forecasts are influenced by many internal and external factors; data, models, analysts, forecasting techniques, reporting delays, production delays, budgets, internal politics, executive agendas, external factors, etc.

Our investigation did not reveal any "new" mathematical techniques, however, the prototype method currently used by NPRST to predict near-term (i.e., 1- to 12-month horizon) Navy officer separations is not used elsewhere and is worth reporting.

Navy Officer Force Separations Forecasting Method

An accurate prediction of separations is very critical since it is used to estimate future vacancies in the officer force, thus impacting current and future personnel manning levels. This is especially critical within 12 months since Navy personnel budgets are subject to individual fiscal years that must be managed as efficiently as possible. Reductions in force (RIF), retirement incentives, and recruiting incentives are just a few force-shaping policies that rely on the forecast of separations.

The method for generating the 12-month separation forecast is a weighted combination of forecasts of scheduled near-term voluntary and involuntary separations, e.g., resignations, retirements, and statutory separations with rates derived from historical transactions.

Requirements for a 12-month forecast of separations are:

- **Projected monthly separations for the next 12 months.** The primary purpose is to provide an accurate estimate of the number, cause, and timing of officers' separations.
- **Uses a simple approach for generating forecasts.** This is especially critical since the officer personnel planners are typically in their job, as a personnel planner, for less than 36 months. Also, personnel planners may or may not have academic training in an analytic discipline, such as, operations research, statistics, mathematics, etc. or previous related work experience.
- **New forecasts must surpass the accuracy of current forecasting methods.** This is the pass/fail criterion for acceptance of a new method of forecasting separations.

Since the Navy maintains comprehensive records of personnel transactions, computing rates using historical data is straightforward, particularly when using the most recent 12-months. The separation rate for a particular category is equal to the total separations for the category and time period divided by the beginning inventory for the same time period, where category refers to the type of separation and time period refers to the time horizon. Initially, the separation rates are calculated using a time period of one fiscal year. Due to significant seasonal trends in officer separations by category, a historical monthly distribution is calculated for each category and applied to the annual separation rates. Separation rates may also be computed by using a weighted average of inventories and separations from different years; however, for this general explanation focus is on a straightforward example using only the most recent 12 months.

Historical Separation Rate Equation

Separation Rate [Resignations, Time Period] =
Total Separations [Resignations, Time Period]/Begin Inventory [Begin of Time Period]
Separation Rate [Resignations, FY94] =
Total Separations [Resignations, FY94]/Begin Inventory [Begin FY94]

Example: .32 = 32/100

Probability of Separation

Probabilities of future separations are derived from records that have a future estimated separation date indicated within an individual's personnel record. Navy officers are required to submit a separation request several months in advance of their desired separation date. This provides an ideal opportunity to exploit the predictive nature of such information. The Navy's master personnel file lists the anticipated separation date for those officers who have submitted separation requests. Probabilities are developed based on the period of time the individual is noted as an "expected" separation.

Using a minimum of 36 months of historical data, individual records are grouped by the number of months until estimated separation and particular separation category. By distinguishing the earliest 24 months from the most recent 24 months (with a 12-month overlap in the middle), a table is created with the groups of individuals having separation indicators relative to the amount of time until expected separation by category. The final 24 months are used to identify the actual separations that occurred and when, for each group of estimated separations in the prior step. For each group of estimated separations by months until separation and type of separation, a distribution of actual months to separation will be calculated. The final step is to compute probabilities using the total number of expected separations, relative to the number of months until expected separation and separation category as the denominator and the actual number of separations (for months until separation 1 through 12), relative to the number of months to expected separation and separation category, as the numerator.

Probability Equation

Probability [Category, # Months to Expected Separation, # Months of Actual Separation] = Actual Separations [Category, # Months to Expected Separation, # Months of Actual Separation]/Expected Separations [Category, # Months to Expected Separation]

Example: Probability [Resignations, 5-months, 6-months] = Actual Separations [Resignations, 5-months, 6-months]/Expected Separations [Resignations, 5-months]

.68 = 68/100

Combining Historical Rates with Near-Term Probabilities

The novelty of this forecasting method is that it allows for blending of projected separations derived from separation requests with projected separations using historical rates. The two forecasts are combined using a "weight" factor. The weight for a projection period determines how much confidence is given to projected separations derived from near-term probabilities against how much is given to historical rates. This is best understood through an example: if the weight factor for month 6 (where month 6 is 6 months away from the current month) is .719, then the separations derived from separation request data is assigned a weight of .719 and the historical separation projection is assigned a weight of .281, derived from 1 - .719 (100% minus the weight assigned to the probability projection). The probability weight factor is derived from the expected accuracy of the probabilities derived from the separation request data, during the analysis of historical data (described above). In this example, there are 100 expected separations, based on records having a separation indicator, for month 6. Stopping at this step would not be sufficient since additional separations occur that are truly "unexpected" and are not tracked. The total projected separations using probabilities is 100 divided by .719 which equals 139 separations. Continuing the example, suppose the projected separations using historical rates, solely, is 200—this appears to represent a situation where separations in the current year will be less than the previous year. The final separation projection for month 6 is 156, derived from (139 * .719 + 200 * .281).

Equation for Combining Projections

*Projected Separations = [# Separations by Probability * Weight + # Separations by Historical Rate * (1 – Weight)]*

Example: 156 = [139 * .719 + 200 * (1 - .719)]

Forecasting and Analysis

The following sections represent the findings from an empirical analysis of Navy civilian workforce data.

Data Collection and Analysis

Historical data used in this study were obtained from the Defense Manpower Data Center (DMDC). It represents Navy civilian personnel transactions that occurred between September 1993 and September 2003 (i.e., fiscal years 1994 to 2003). The data were received as ten text files, one for each fiscal year, which included a summary of all transactions that occurred during the fiscal year. For example, the September 1994 file contains all recorded transactions for any Navy civilian between October 1, 1993 and September 30, 1994. Each file was converted into Microsoft Excel, Microsoft Access, and SPSS for further processing and analysis. Several macro programs were written to manipulate and analyze the data (e.g., payplan and occupational code (series) fields were used to produce a community code field), to specify the occupational group for each individual record. Several other fields were produced to perform regression analysis.

Two significant changes to the data processing which effects the transaction files are noted; as of fiscal year 1997 service computation date was reduced to six characters CCYYMM, where CC represents century, YY represents year, and MM represents month. In previous years service computation date was represented using eight characters to include DD as record of the actual date (i.e., CCYYMMDD). Nature of action code value 894 (pay adjustment) was excluded after fiscal year 2001. The nature of action field was useful in data validation because every employee received a cost of living pay adjustment each year. Without tracking pay adjustment there is no assurance all personnel are included in the transaction files; it is virtually impossible to produce a master file from annual monthly transactions. NPRST has requested the monthly master files from DMDC, including the identical fields as received in the transaction files. Unfortunately, the master files were not received prior to completion of this report.

Table 1 represents the fields that were requested versus the fields that were received. Two additional fields for community code and involuntary and voluntary separation code were added to develop forecast models. Table 2 lists the civilian community definitions.

Table 1
Community code definitions

Data Element Name (Requested)	DCPDS Table	DMDC (provided)
Accession/Separation Code	120/490	SSN
Acquisition Corps Identifier	211	Service Computation Date
Additional Legal Authority	463	Citizenship
Age in Months	N/A	Work Schedule
Citizenship	N/A	Effective Date of Personnel Action
Date Entered Grade	N/A	Occupation Code (Series)
Disabled Indicator	N/A	Handicap Code
Drawdown Indicator	616	Pay Basis
Education Level	469	Agency
Employee Category	N/A	Bureau
Grade	N/A	PayPlan
Minority Group Designator	132	Grade
Mobilization Position Indicator	472	Education Level
Nature of Action Code	85	Race
Organization Code	Activity Level Table	Length of Service in Months
PayPlan	484	Age
Primary Legal Authority	463	Date of Current Grade
Reserve Category	123	Yearly Compensation (Salary)
Retirement Code	101	Tenure
Salary Annual	N/A	Nature of Action Code
Series	466	Pay Rate
Service Computation Date (Leave)	N/A	Retirement Eligibility
Service in Months	N/A	Overseas Emergency Essential
Social Security Number	N/A	Retirement Code
Special Program Indicator	N/A	Legal Authority 1
Type Command	N/A	Legal Authority 2
Type of Employment	163	Community Code *
Voluntary Separation Incentive Pay Indicator	N/A	Voluntary/Involuntary Separation Code *

* This field was produced by NPRST.

Table 2
Civilian community definitions

CC*	Community	Series
01	Administration	301, 302, 303, 305, 309, 312, 313, 318, 319, 322, 326, 341, 342, 343, 344, 350, 351, 356, 357, 382
02	Analyst	110, 130, 131, 1515
03	Community Support	030, 180, 185, 186, 187, 188, 189, 1173
04	Contracts	1101, 1102, 1105, 1106, 1160
05	Education & Training	1701, 1702, 1710, 1720, 1750
06	Environment	0028, 0029, 819
07	Facilities	0020, 0021, 803, 807, 808, 809, 810, 817, 818, 828, 1103, 1104, 1107, 1130, 1144, 1150, 1170, 1171, 1176, 1601, 1640, 1658, 1667, 1670
08	Financial	510, 503, 505, 510, 511, 525, 530, 540, 545, 560, 561
09	HR	201, 203, 260, 361, 140, 142
10	Industrial Trades	>=2500
11	Intelligence	132, 134
12	IT/IM	2210, 1550, 332, 335, 391, 1411, 1412, 1420, 1421
13	Legal	904, 905, 950, 962, 963, 986, 998, 998, 990, 992
14	Logistics	346, 2001, 2003, 2005, 2010, 2030, 2032, 2050, 2091, 2101, 2102, 2130, 2131, 2135, 2144, 2150, 2151, 2152, 2154, 2161, 2181
15	Manufacturing	806, 894, 895, 896, 1152, 1910, 1654
16	Media & Public Relations	0170, 0193, 1001, 1008, 1010, 1015, 1016, 1020, 1035, 1040, 1046, 1051, 1060, 1071, 1082, 1083, 1084, 1087
17	Medical	405, 601, 602, 603, 610, 620, 621, 622, 630, 631, 633, 636, 638, 640, 644, 645, 646, 647, 648, 649, 651, 660, 661, 662, 664, 665, 669, 670, 671, 673, 675, 679, 681, 682, 0683, 688, 690, 698
18	Program Management	340
19	Safety	0018, 0019
20	Science & Engineering	101, 102, 150, 401, 403, 404, 405, 408, 410, 413, 414, 457, 462, 486, 487, 493, 802, 804, 830, 840, 850, 855, 856, 858, 861, 871, 873, 880, 881, 890, 892, 893, 1301, 1310, 1411, 1313, 1315, 1320, 1321, 1330, 1340, 1350, 1360, 1361, 1371, 1374, 1397, 1520, 1521, 1529, 1530, 1531
21	Security & Law Enforcement	006, 072, 080, 081, 083, 085, 086, 1801, 1802, 1810, 1811, 1812, 1815,

* Community Code

Forecasting and Analysis Results

Forecasting capability was evaluated for three methods; naïve, time-series, and regression. Resignations and retirements were evaluated using the Naïve and Time-series methods, while the Regression method was used to evaluate retirements only – since N11's primary concern is with the increasingly large population of retirement eligible employees.

Historical personnel inventories for general schedule (GS) and wage grade (WG) are represented below (see Figure 1).

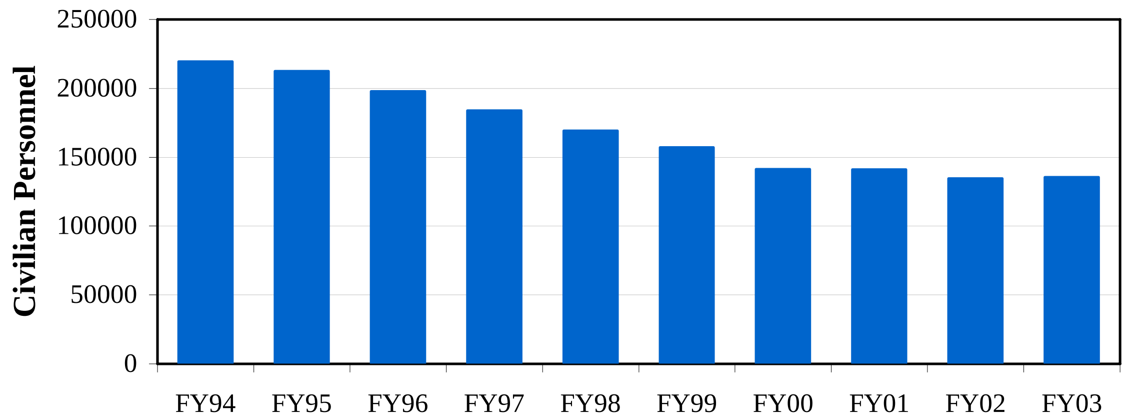

Figure 1. General Schedule & Wage Grade historical totals.

Naïve and Time-series Results

The Administration, Industrial Trades, and Science and Engineering communities (occupational groups) were selected as a representative sample to test forecasting capabilities based on historical civilian personnel behavior. These three groups represent clerical, industrial and professional staff members as well as occupational size ranges. See Appendix B for a graphical display of historical patterns of resignations and retirements.

The Naïve Forecast method is defined as assuming next year's forecast will be the same as the current year's actual, or next year's rate of change will be the same as the current year's rate of change. This could also be considered last year's actuals as the current year's forecast. Time-series Forecast method assumes a pattern will continue into the future. The accuracy of each method will improve based on interaction from the end-user by determining whether to adjust forecast rates given knowledge of historical events, and/or knowledge of current internal and external factors. The assumptions presented in this report are intentionally simple to provide an objective evaluation. For example, the Time-series method uses an even weighting of the available historical rates, while in a real-world scenario the analyst is likely to choose a different weighting scheme for various reasons.

The Naïve Forecast method was evaluated using the previous year's actuals and rate of change. For the Time-series Forecast method, the rates of change for each previous year were averaged (i.e., to predict fiscal year 1996 resignations, the actual rates of change from fiscal year 1994 and 1995 were averaged). Annual resignation and retirement forecasts error results are displayed graphically in Appendix C. Since individuals must meet eligibility requirements before retiring, the aggregated retirement eligibility numbers are provided in Appendix B. Eligibility was determined by combining information in the retirement eligible code and the retirement code to determine if an individual qualified for a full or reduced annuity, where reduced annuity is considered an early retirement.

Tables 3 and 4 summarize the mean prediction accuracy for the Naïve and Time Series methods over the 10-year period considered. Table 3 represents the accuracy in predicting resignations, which was not modeled via the regression method. The Naïve Rate method was a consistently more accurate method for predicting resignations. Its mean accuracy range was 80.43% (Science & Engineering) to 94.76% (Administration).

Table 3
Resignations

Occupational Group	Method	Mean Prediction Accuracy
Administration	Naïve (Actual)	89.11%
	Naïve_Rate	94.76%
	Time_Series	85.88%
Industrial Trades	Naïve (Actual)	76.03%
	Naïve_Rate	84.89%
	Time_Series	56.68%
Science & Engineering	Naïve (Actual)	76.95%
	Naïve_Rate	80.43%
	Time_Series	70.11%

Table 4 represents the accuracy in predicting retirements using the Naïve and Time Series methods. Regression model accuracy is presented in Table 5. The greatest mean prediction accuracy differed by occupational series (i.e., Time Series for Administration, 84.14%; Naïve (Actuals) for Industrial Trades, 73.49%; and Naïve Rate for Science and Engineering, 82.59%). The Time Series method performed poorly for the Industrial Trades occupation, which can be attributed to the sharp decrease in retirement rates for this group between fiscal years 1995 and 1998 (see Figure B-10).

Table 4
Retirements

Occupational Group	Method	Mean Prediction Accuracy
Administration	Naïve (Actual)	81.89%
	Naïve_Rate	71.34%
	Time_Series	84.14%
Industrial Trades	Naïve (Actual)	73.49%
	Naïve_Rate	72.80%
	Time_Series	0.00%*
Science & Engineering	Naïve (Actual)	79.74%
	Naïve_Rate	82.59%
	Time_Series	57.67%

* Error percentages greater than 100% have no validity.

Regression Model Results

A logistic regression model was developed to evaluate retirement behavior for the selected occupational groups (e.g., Administration, Industrial Trades, and Science & Engineering). The objective of the regression model is to determine the variable(s) with statistical significance and forecast accuracy of the model (see Appendix E for the logistic model and estimates; see Appendix D for a description of explanatory variables).

The binary logistic model was estimated based on the probability of an individual not retiring when eligible. The response level was coded as one for individuals who retired and zero for individuals who did not retire. GS-14 to GS-15, white employee within the education 5 category (Post Bachelor's, Post Master's, or Post Doctorate) was used for the Y-intercept (base case) (see Appendix E). Detailed frequency values for each variable are provided in Appendix F.

Table 5
Regression Model Accuracy

	Actual Observations	Actual Retired	% Retired	Prediction Accuracy
Admin	17012	2014	12%	71%
Indus. Trades	34565	3797	11%	75%
Sci. & Eng.	14000	1557	11%	72%

Although there is a wide range of observations per occupational group, prediction accuracy of the regression model was fairly consistent across occupational groups. As compared to other methods (see Table 4), the regression model predicted with the greatest accuracy for the Industrial Trades group, which underwent the most drastic change in retirements (i.e., 3000 retirements in fiscal year 1995 to 1000 retirements in 1998). Administration and Science and Engineering groups changed incrementally from year to year with high prediction accuracy using Naïve and Time Series methods.

The prediction accuracy of the regression model is expected to improve with inclusion of additional data (i.e., greater number of observations which could only be achieved with more historical data, separating retirement behavior by retirement program (CSRS, FERS), wealth information, number of college age dependents, retirement indicators, and other.).

Prototype Model Specifications

The ultimate product will be a prototype workforce planning tool capable of aiding analysts in making decisions for the Navy's civilian workforce.

Community managers will need to communicate ideas and findings related to their civilian communities to N11 leadership, community leaders, and members within the community. N11 leadership has the authority to execute policy changes. Community leaders have the knowledge and closeness with personnel and daily operations within the various communities. Individual members are likely to have questions and concerns related to their individual careers.

Several factors contribute to specifying the prototype model: technical aptitude of users, expected frequency of use, and availability of underlying data.

Excluding exceptions such as analyst or engineering and science communities, individuals serving as community managers are less likely to have sufficient analytical training or experience. It is very critical to provide a tool that is suitable for the majority of users who lack analytic knowledge and experience. The interface should make mathematical formulas and calculations transparent. However, each underlying methodology should have a complete explanation through a help menu. For example, if the tool is being used to create a forecast of future retirements for a community, the tool should allow the user to select among available prediction methods; however, the tool should have a default method. The motivation is to ensure users' trust and arm them to defend the results.

Daily use of the model for creating workforce plans is not anticipated; however, there is likely to be daily use for accessing underlying data and trend analysis once the tool is in use and accepted as credible. Workforce plans are likely to be created and/or revised quarterly, bi-annually, or annually. Expect workforce plans to be briefed and questioned frequency—community managers will need to execute alternative runs and evaluate prior runs on a continuous basis. Quarterly updates to underlying data should be sufficient given the stability of the civilian workforce. Ideally, the quarterly updates will

capture changes to personnel records recorded in the nature of action code field. Annual revisions to the rates of change are recommended, particularly focusing on accessions, separations, personnel actions, and policies affecting career ladders.

The user interface must be carefully designed to ensure intuitiveness. Recommend the user interface be modeled after Microsoft compatible programs where the file options are accessible from the top of the screen, following the default order (e.g., file, edit, view, etc.). If the design includes a tree structure, the tree should be positioned along the left side of the screen. Where applicable, text-based help information should be displayed by placing the pointer over a button, pull-down, or other option. A standard processing order should be maintained for each individual community being evaluated. Functionally, recommend the user have access to the most current data for viewing. This would represent a starting inventory. The standard processing order is computation of separations, accessions, other personnel actions (e.g., promotion, skill designation changes, etc.), and finally the resulting inventory for future time horizons.

Conclusions and Recommendations

Recommend developing a prototype model because: (1) prototypes allow evaluation on a small scale prior to full development, (2) prototypes are less expensive to develop than full systems, (3) prototypes are scaled small enough to allow for continuous development, and (4) prototype development is more suitable where user involvement is possible. An intuitive user interface is paramount, and will decrease the time necessary for new community managers to become comfortable. It will also enhance the likelihood it will be used.

Although CIVFORS is not a direct fit for the Navy's civilian workforce, we recommend the developers of the Navy's civilian workforce planning tool(s) use CIVFORS as a guide. CIVFORS has incorporated much of the functionality the Navy will need to duplicate for planning its workforce, primarily the sub-modules and use of multiple forecasting methods.

Recommend Naïve and Time Series methods as baseline methods because they are simple to implement and interpret, and they have high prediction accuracy when changes in the workforce are slight. A regression model is strongly recommended for use as a prediction method in more volatile environments, such as pronounced changes in workforce turnover.

Appendix A:
Forecasting Glossary

Forecasting Glossary

Aggregate Forecast – Forecast of an organization as a whole.

Auto-Regression or Auto-Regressive Process – Where results from one period are regressed on the previous period.

Auto-Regressive Moving Average (ARMA) Process (Model) – Where the auto-regressive and moving average processes are combined. It is often called ARMA mode or the Box-Jenkins Model.

Autocorrelation – Correlation within a series. For example results from 1996 are related to results from 1995, and results from 1995 are related to results from 1994.

Auto-correlated Time Series – A time series in which the current value of a series depends on the past value.

Back Forecasting – Making forecasts of periods for which actuals are known. Also known as ex-post forecasts.

Base Period – A period in time from which comparisons of other time periods are made.

Best Linear Unbiased Estimator (BLUE) – The criterion used in regression modeling to select the best estimator form a number of Unbiased Estimators.

Bias – It is often referred to an error resulting from an error in data gathering, faulty program design, mistakes on the part of personnel, or data sources.

Bottom-up Forecasting – Forecasts that originate from the bottom. For example, obtain inventory forecasts from community managers of different communities and then add them together to arrive at the aggregate forecast.

Box-Jenkins Model – A time-series model named after the developers of this model. It combines the Auto-Regressive Process with a Moving Average.

Bullwhip Effect – In case of stock out, customer tends to order more than it needs which corrupts the real pattern.

Categorical Variable – A qualitative variable created by classifying observations into categories. For example, incomes could be classified into categorical variables, low, medium, and high based on specific ranges of income levels.

Casual Model – A model that assumes that the variable to be forecast exhibits a cause-and-effect relationship with one or more other variables. Regression/econometric models are causal models.

Census – A complete enumeration of the universe (population). In contrast, sample is a portion of the universe.

Classical Decomposition – A time series model, which decomposes the data into trend, cycle, seasonality and randomness.

Coefficient Term – It is a slope of the line. It shows how the dependent variable, on the average, changes with a once unit change in the independent variable.

Consensus Forecast – Forecasts, which are jointly agreed upon. Or, average of forecast given by different individuals.

Correlation Coefficient – A standard measure of relationship between a dependent and independent variable.

Customers – Customers of a vendor or distributors, wholesalers and/or retailers.

Cycle – Cyclical fluctuations are those that occur regularly but not periodically. The length of a cycle is always more than one year.

Data Warehouse – Where data is stored.

Delphi – This is a judgmental technique of forecasting where a panel of experts are asked to give their own forecasts which is distilled to arrive at a final forecast.

Demand – Booking orders.

Dependent Demand – It represents the demand of vendor's factory (raw material, etc.), vendor's distribution center demand, which depends on the customer distribution center's demand, customer retail store's demand, which depends on the demand of final consumers.

Dependent Variable – A variable we wish to forecast. In regression analysis the variable being predicted is the dependent variable.

Disaggregated Forecasts – Breaking up the total company forecast into categories, e.g., SKUs.

Durbin-Watson Test – Diagnostic tool used to test a regression model. Its value varies between 1 and 4. The model is the best if its value is 2. Normally, the value is between 1.5 and 2.5 is acceptable.

Econometric Forecasting – Where a model encompasses more than one equation to make a forecast.

Econometric Indicator – It provides an indication of how the economy is behaving.

End-User – Ultimate user of a forecast.

Ex-Ante Forecast – Preparing forecasts for periods for which actuals are not known.

Endogenous Variable – An internal variable which can be changed.

Exogenous Variable – An external variable which is not controlled by the forecaster.

Explanatory Variables – The variables that drive the sales, they are used to predict values of a dependent variable; sometimes called independent variables.

Ex-post – Preparing forecasts for periods for which actuals are known.

EVA – Economic value added.

Fitted Values – The predicted valued derived from a regression model by applying the regression coefficients to the independent variables.

Forecast Horizon – The number of time periods out to be forecasted (i.e., 1 month out, 1 quarter out, 1 year out, etc.)

Forecasting Process – The process outlines who will provide the information used for preparing forecasts; how it is gathered; after information is obtained how it is processed and used for preparing statistical forecasts; and once statistical forecasts are prepared who participate in the process to arrive at consensus forecasts.

F Statistics – In a regression model it is used to determine the overall performance of a model.

Forecast System – Mechanizing the forecasting process including the use of software and hardware.

Forward Buy – Occurs when an account buys extra quantity during the deal period to be sold after the deal has ended.

Independent Demand – Represents consumption demand.

Independent Variables – The variables that drive the sales. They are used to predict values of a dependent variable. They are also called explanatory variables or drivers.

Intermittent Demand – These are the products that have no demand or many months of sporadic demand in other months.

Lead Time – Time needed to make any change in production plan or ordering raw materials. Or, amount of time required to provide (or produce) a product to an inventory location. Or, time needed to make any change in production plans.

Leading Indicator – Economic indicator whose peaks and troughs during the business cycle tend to occur before the general economy. Stock market prices are generally considered as a leading indicator of the economy.

Macro-forecasts – Forecasts of the economy as a whole. For example, forecasts of GDP and employment.

MAPE – Mean Absolute Error, the average percent error with signs ignored.

Matured Products – Products that have passed their growth stage in terms of demand.

Micro-forecasts – Company level forecasts. For example, sales forecast.

Multicollinearity – When two independent variables are highly associated (correlated) with each other. It is not considered good in regression modeling.

MSE – Mean squared error. Here errors are first squared and then their average is computed.

Naïve Forecast – Next year forecast is the same as the current year actual.

Observations – Number of periods used in a forecasting model.

Operational Forecasts – Short-term forecasts, usually of less than one year.

Outlier – A value that is unusually too large or too small.

Pooling Effect – When a consumer is not able to find the actual size desired, a high probability exists that the consumer may purchase another size in the same product family prior to switching to a competing brand.

Phasing – Percent of annual sales realized in a given month.

Price Elasticity – How sensitive is the sales to price. Highly elastic if a small change in price leads to a large change in demand. Highly inelastic, is a large change in price leads to a small change in demand.

Product Life Cycle – Refers to a life cycle of a product. The product forms a S curve with four stages of development – introduction, growth, maturity, and decline.

Qualitative Forecasting – Refers to judgmental approach to forecasting.

Quantitative Forecasting – Refers to statistical approach to forecasting.

Regression – It is a causal method of forecasting which assumes that the variable to be forecasted exhibits a cause/effect relationship with one or more variables (factors).

Residual – It is equivalent to a forecast error – the actual minus the fitted forecast value.

Safety Stocks – Buffer stock used to compensate for uncertainties in demand during lead-time.

Scenario Forecasting – A judgment technique of forecasting where several set of circumstances are constructed which form boundaries within which the actual number is expected to lie.

Seasonality – Seasonal fluctuations are those occurring regularly and periodically and the length of a cycle is always less than one year.

Sell in Forecast – Forecast of shipment from the manufacturer to retailer.

Sell Through Forecast – Forecast of sales to end-use consumers.

Shipping Data – Data of merchandise shipped.

Spatial Autocorrelation – Often arises in a cross sectional data where a change in one region may cause a change in the activity in other region because of close economic linkages.

Standard Deviation – Measure of variations within a series. For example, how errors vary over different periods.

Stock-Out – When inventory is not available to meet orders in a timely manner.

Strategic Forecasts – Long-term forecasts, usually of more than one year.

Tactical Forecasts – Short-term forecasts, usually less than one year. Also called operational forecasting.

Time Series Models – Where it is assumed that past pattern will continue in the future. Here one needs only data of series to be forecasted.

Top-Down Forecasting – Here the forecast is first prepared of the company as a whole, which is then disaggregated into category level forecasts.

T Test – It is used to determine in a regression model whether the impact of a certain independent variable is significant or not. In other words, whether we should keep the variable in or throw it out.

Trend – It is statistically computed. It show how, on the average, sales is moving, upward or downward.

Unconstrained Demand – What could have been sold if there were no problem in production or anything else, which might have cut down the sales.

Univariate Models – Here one needs only the data of series to be forecasted. Time series models are univariate models.

Validation – The process of testing whether the model is valid or not.

Appendix B:
Historical Patterns of Resignations and Retirements

Historical Resignation Patterns

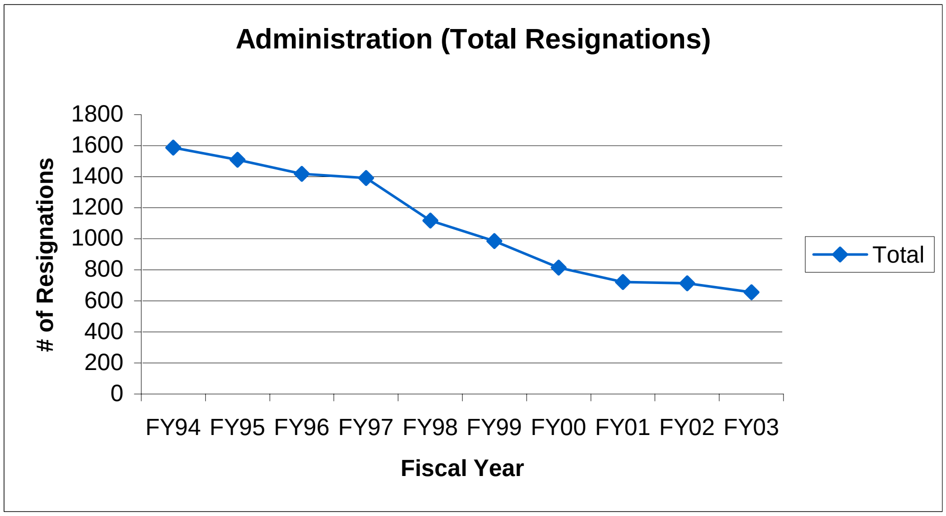

Figure B-1. Total resignations for Administration.

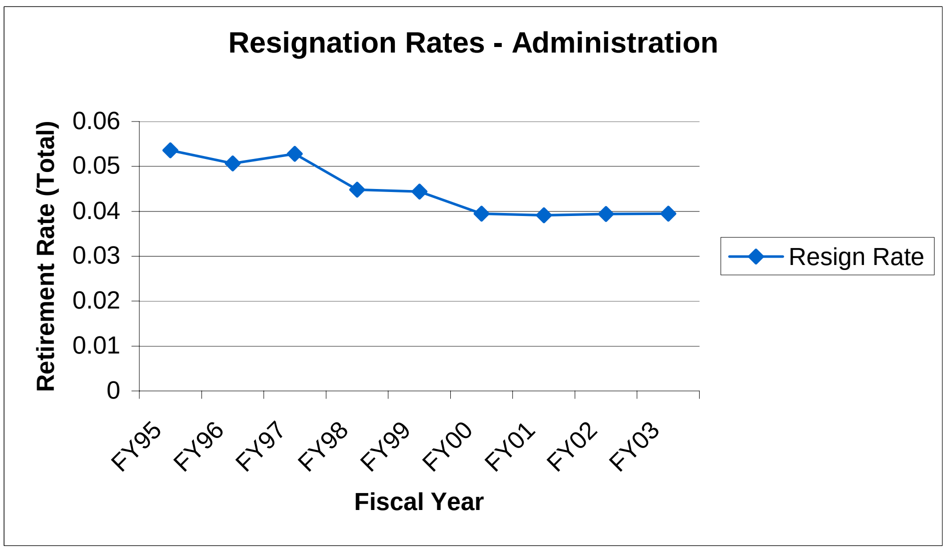

Figure B-2. Resignation rates for Administration.

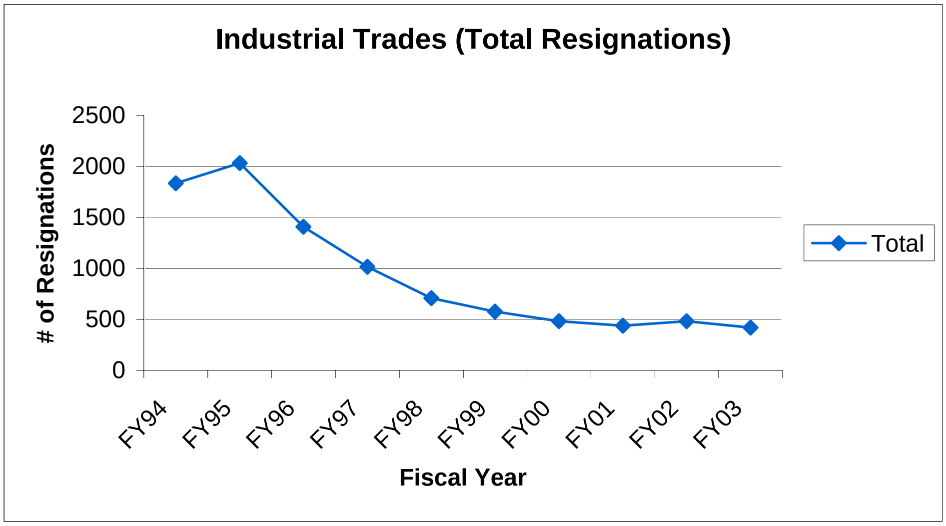

Figure B-3. Total resignations for Industrial Trades.

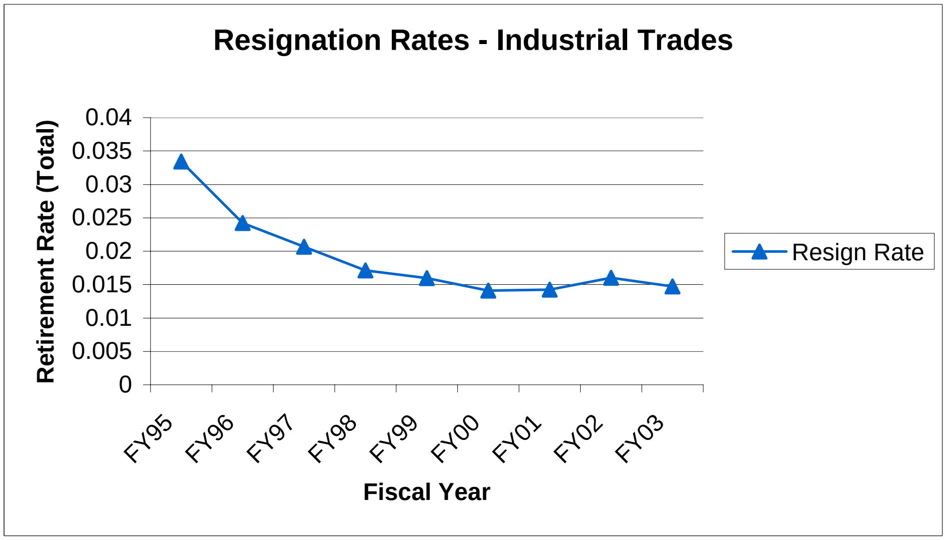

Figure B-4. Resignation rates for Industrial Trades.

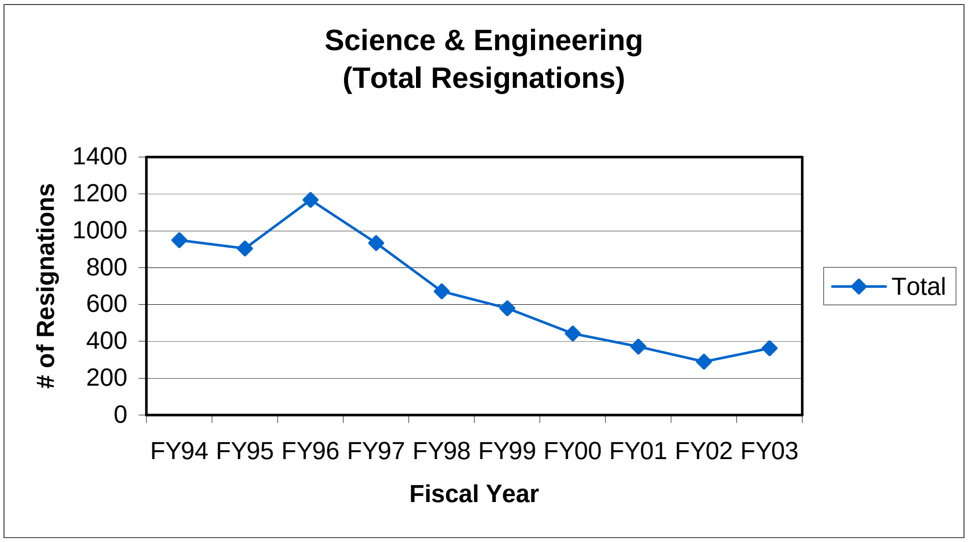

Figure B-5. Total resignations for Science and Engineering.

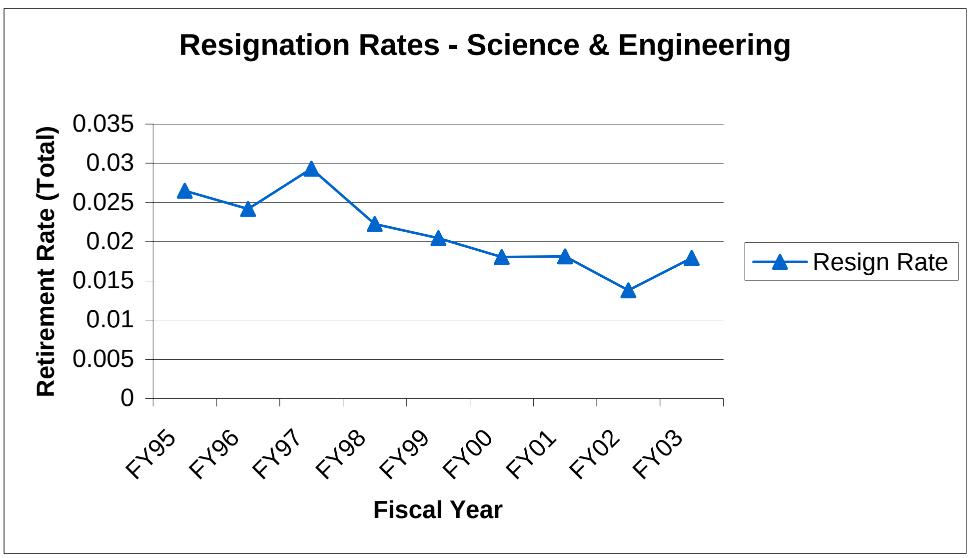

Figure B-6. Resignation rates for Science & Engineering.

Historical Retirement Patterns

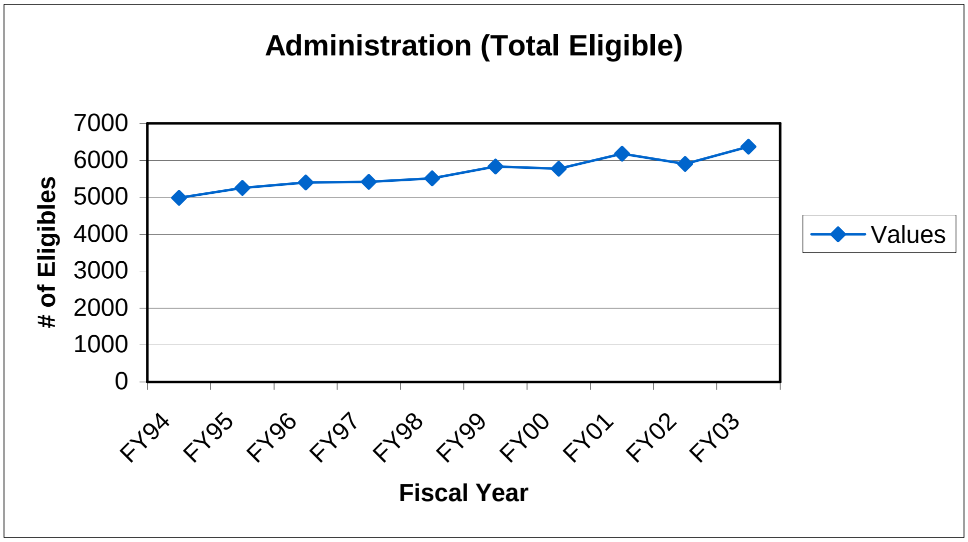

Figure B-7. Total eligible for Administration.

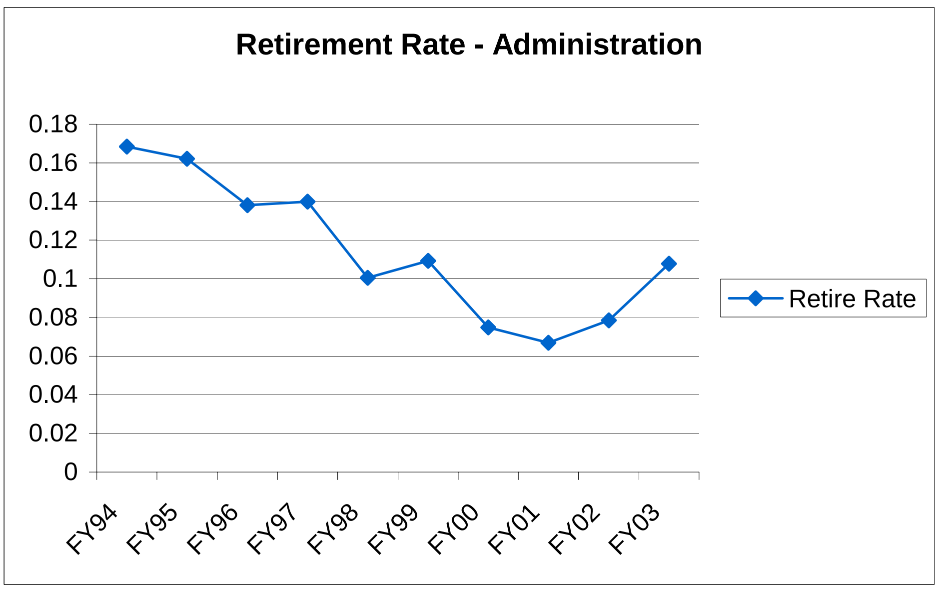

Figure B-8. Retirement rate of eligible Administration

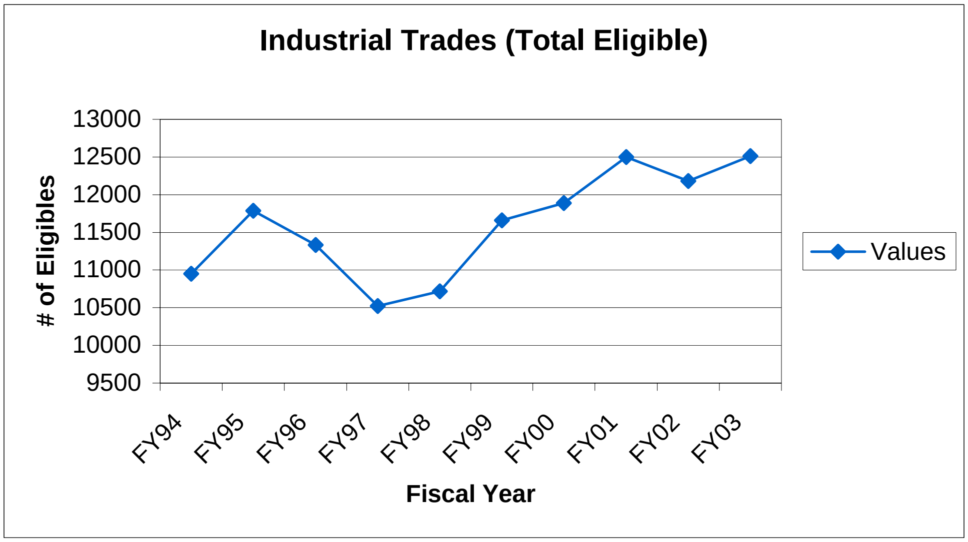

Figure B-9. Total eligible for Industrial Trades.

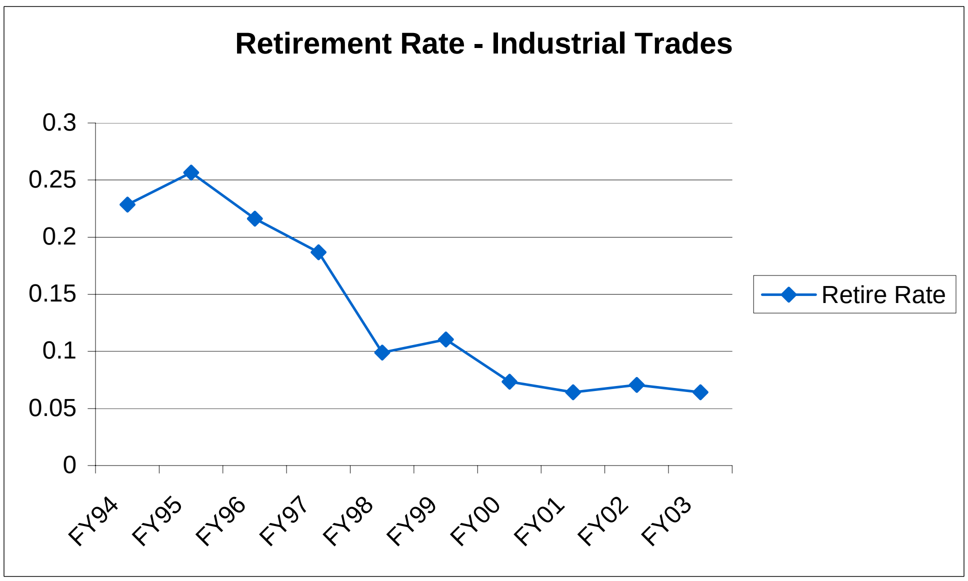

Figure B-10. Retirement rate of eligible Industrial Trades

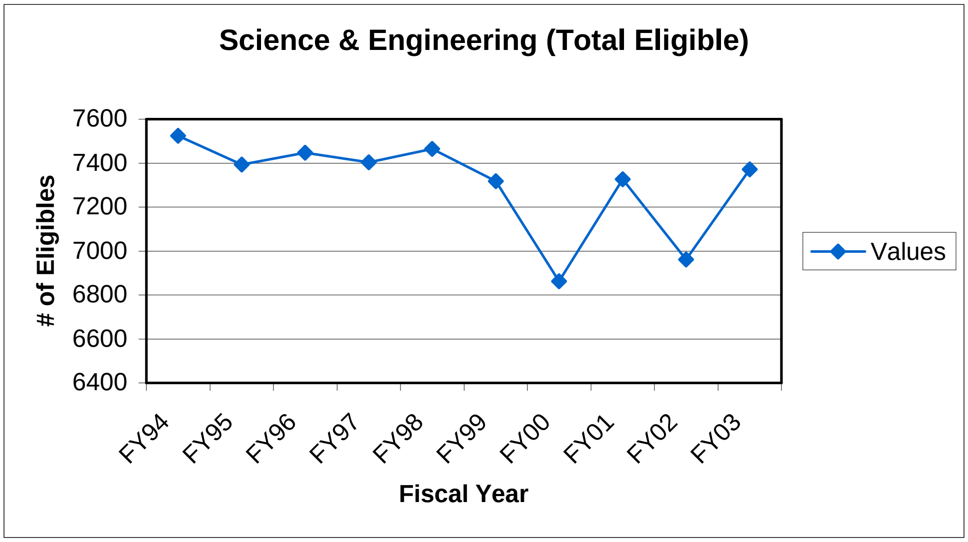

Figure B-11. Total eligible for Science and Engineering.

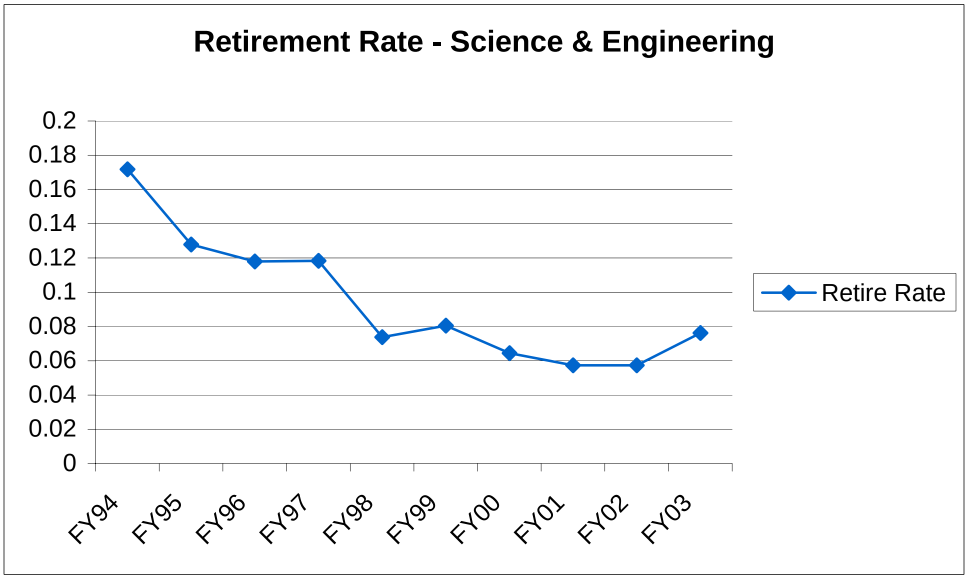

Figure B-12. Retirement rate of eligible of Science and Engineering.

Appendix C:
Resignations and Retirements Forecast Results

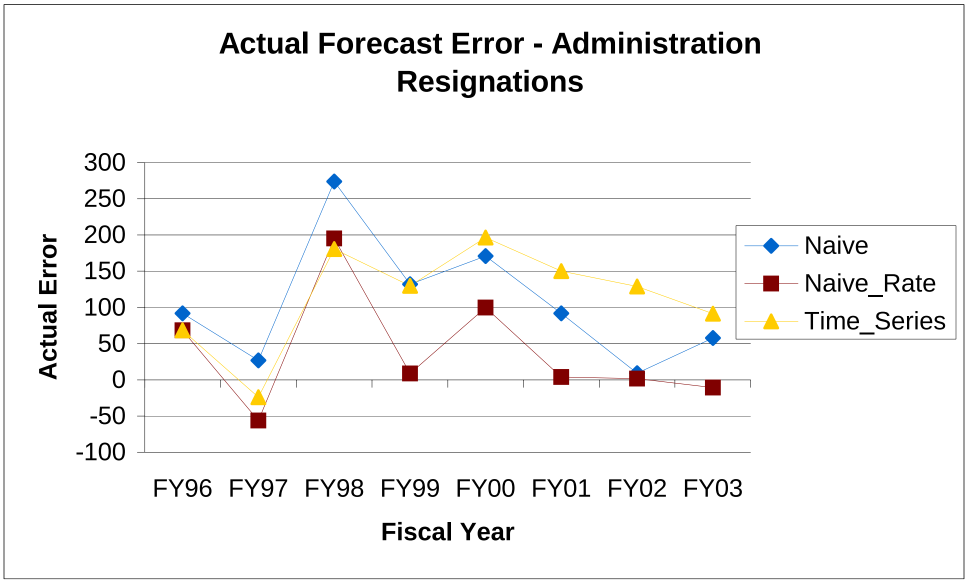

Figure C-1. Forecast minus Observed Resignations

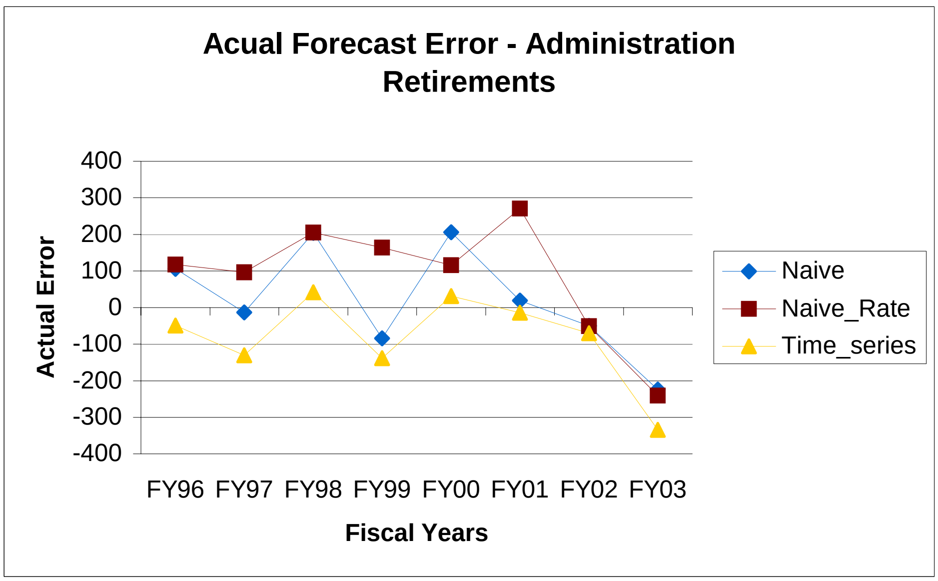

Figure C-2. Forecast minus Observed Retirements

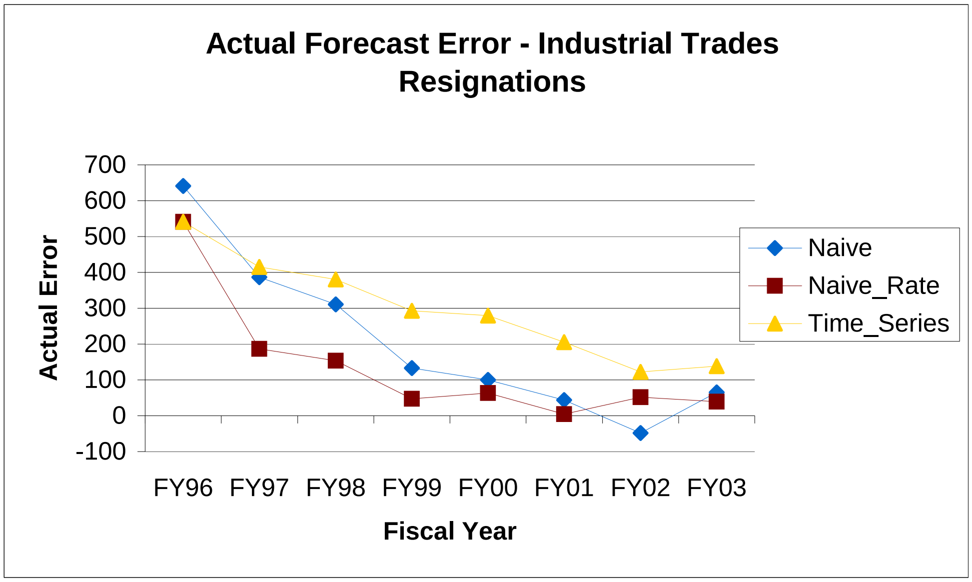

Figure C-3. Forecast minus Observed Resignations

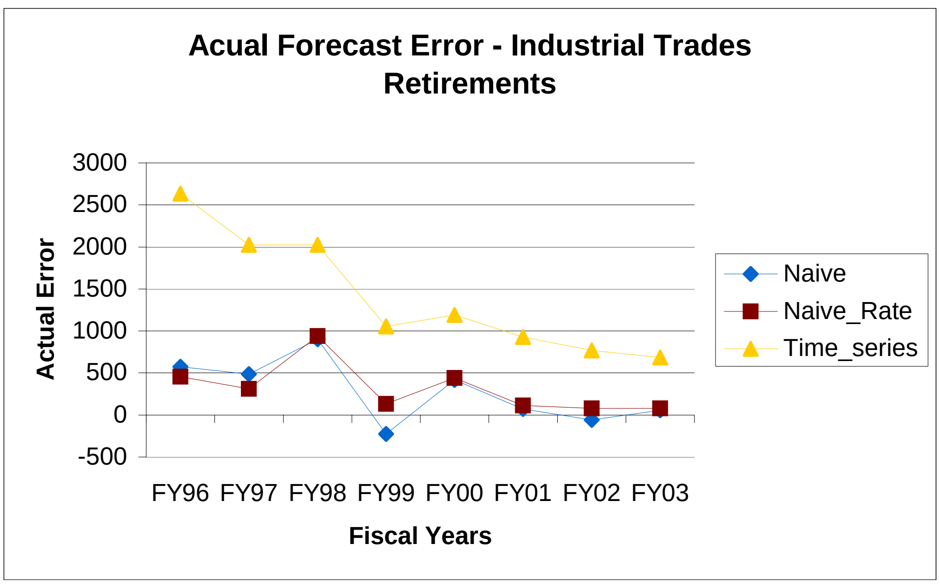

Figure C-4. Forecast minus Observed Retirements

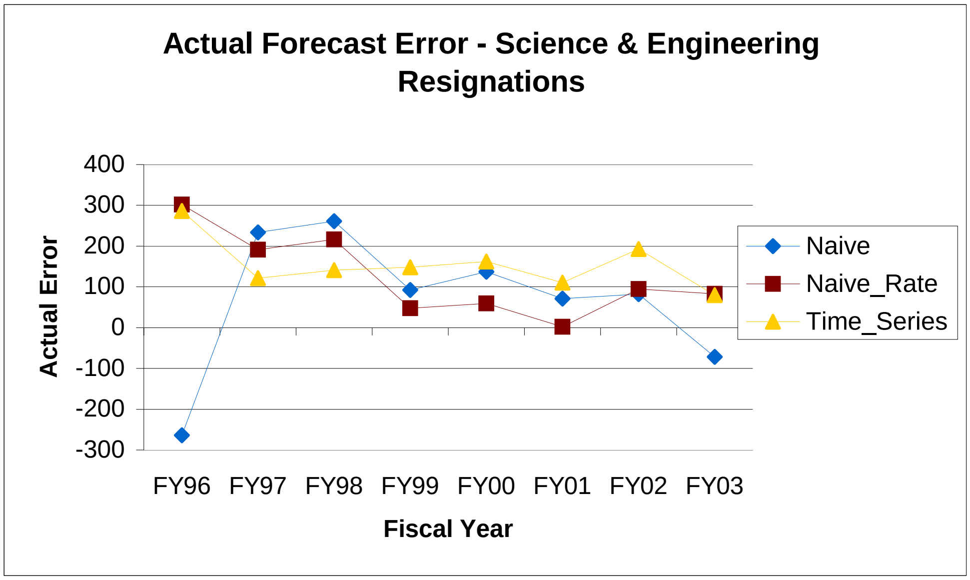

Figure C-5. Forecast minus Observed Retirements

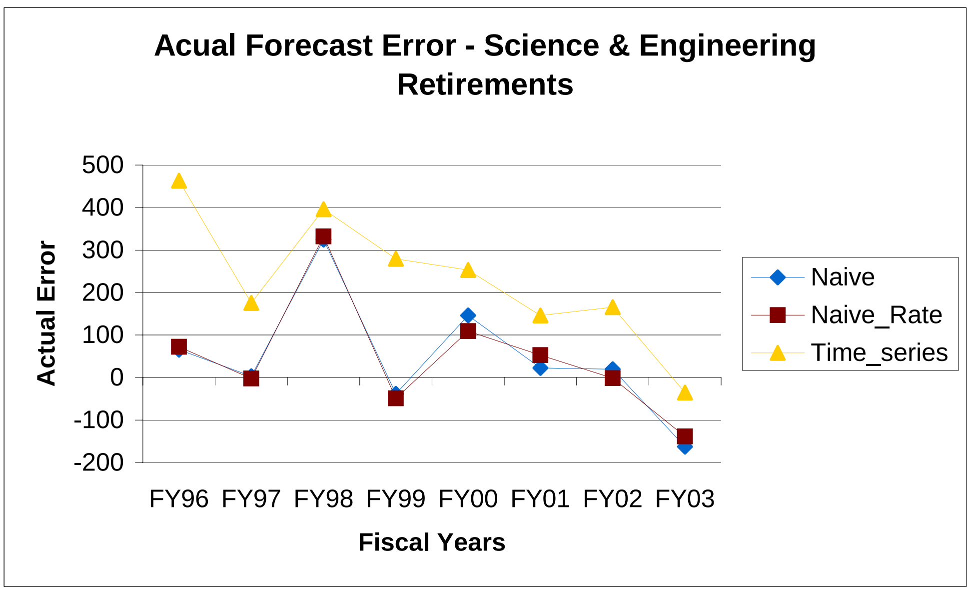

Figure C-6. Forecast minus Observed Retirements

Appendix D:
Explanatory Variables

Table D-1
Race (Minority Group)

Code	Minority Group	Data Element Code
Race 1	Non-Hispanic Black	1
Race 2	Hispanic	2
Race 3	American Indian/Alaskan Native	3
Race 4	Asian/Pacific Islander & Hawaiian Asian	4
Race 5	White Non-Hispanic	7

Table D-2
Education Codes

Education	Description	Date Element Code
Edu 1	Some elementary school/completed elementary/some high school (did not complete)	1, 2, or 3
Edu 2	High School Graduate (or certificate of equivalency	4
Edu 3	Terminal Occupation Program (did not complete)/Terminal Occupation Program (certificate of completion–diploma or equivalent)/some college (< 1 year)/1 year of college/2 years of college/associate degree/3 years of college/4 years of college	5, 6, 7, 8, 9, 10, 11, or 12
Edu 4	Bachelor's Degree	13
Edu 5	Post Bachelor's/1st Professional Degree/Master's Degree/Post Master's Degree/6 Year Program/Post 6 Year/Doctorate Degree/Post Doctorate	14, 15, 16, 17, 18, 19, 20, 21, or 22

Appendix E:
Regression Model Estimates

Regression Model Estimates

A logistic regression is a mathematical modeling approach that can be used to describe the relationship of several X values to a dependent variable. The model's general functional form is

Equation 1 $P(X) = 1/1 + \exp[-(\alpha + \sum \beta i X i)]$.

The following logistic model was estimated:

Equation 2 $P(Y) = 1/1 + \exp[-(\alpha + \sum \beta i X i)]$ (Note: Intercept value equates to α, and X_i is 0 or 1.)

Table E-1
Administration estimates

Parameter	Estimate (βi)	Error
Intercept (α)	6.6465*	0.3706
Race 1	0.1281	0.0882
Race 2	0.1388	0.1498
Race 3	0.1341	0.2664
Race 4	0.3513	0.1172
Edu 1	-0.4839	0.2369
Edu 2	-0.5423*	0.1289
Edu 3	-0.6092*	0.1263
Edu 4	-0.3266	0.1436
GS1_5	-3.1034*	0.2717
GS6_9	-2.6896*	0.2502
GS10_13	-1.6371*	0.1970
Age 1	3.8317*	0.1229
Age 2	1.3791*	0.0565
Salary	-0.00005*	3.677E –6
LOS	-0.00459*	0.000280

*Statistically significant at α < 0.05.

Table E-2
Industrial Trades Estimates

Parameter	Estimate (βi)	Error
Intercept (α)	6.1233*	0.4641
Race 1	0.0575	0.0532
Race 2	0.2217	0.0919
Race 3	0.0811	0.1999
Race 4	0.5644*	0.0585
Edu 1	-0.5994	0.3802
Edu 2	-0.5812	0.3775
Edu 3	-0.6596	0.3782
Edu 4	-0.2881	0.4074
GS1_5	-1.6259*	0.2276
GS6_9	-1.2037*	0.2250
GS10_13	-0.7403	0.2219
Age 1	3.5154*	0.0651
Age 2	1.6429*	0.0426
Salary	-0.00007*	3.189E –6
LOS	-0.00528*	0.000229

*Statistically significant at $\alpha < 0.05$.

Table E-3
Science & Engineering Estimates

Parameter	Estimate(βi)	Error
Intercept (α)	7.5171*	0.3472
Race 1	-0.0160	0.1681
Race 2	0.1298	0.1818
Race 3	-0.0890	0.3767
Race 4	0.5675*	0.1222
Edu 1	-0.4016	0.3290
Edu 2	-0.8267*	0.1074
Edu 3	-0.7419*	0.1033
Edu 4	-0.2396	0.0969
GS1_5	-2.2878*	0.1957
GS6_9	-2.3247*	0.2282
GS10_13	-1.7593*	0.1688
Age 1	4.3142*	0.1586
Age 2	1.2991*	0.0569
Salary	-0.00006*	3.583E –6
LOS	-0.00435*	0.000314

*Statistically significant at $\alpha < 0.05$.

Appendix F:
Explanatory Variable Frequencies

Table F-1
Continuous Variables

	Administration		Industrial Trades		Science & Engineering	
	Salary ($)	**LOS (Months)**	**Salary ($)**	**LOS (Months)**	**Salary ($)**	**LOS (Months)**
N	17061	17061	34575	34575	14009	14009
Mean	39431.27	280.22	36704.61	276.86	52463.13	307.42
Std. Dev.	15853.85	82.81	6732.81	73.69	12312.70	80.61
Minimum	0	60	0	60	0	60
Maximum	110682	667	75153	671	110682	667

Table F-2
Administration Discrete Variables

	Frequency	Percentage
Race 1 – Retired	2073	12.19
Race 1 – Non-Retired	14939	87.81
Race 2 – Retired	562	3.30
Race 2 – Non-Retired	16450	96.70
Race 3 – Retired	151	0.89
Race 3 – Non-Retired	16861	99.11
Race 4 – Retired	1026	6.03
Race 4 – Non-Retired	15986	93.97
Race 5 – Retired	13200	77.59
Race 5 – Non-Retired	3812	22.41
Edu 1 – Retired	230	1.35
Edu 1 – Non-Retired	16831	98.65
Edu 2 – Retired	7505	43.99
Edu 2 – Non-Retired	9556	56.01
Edu 3 – Retired	6544	38.36
Edu 3 – Non-Retired	10517	61.64
Edu 4 – Retired	1691	9.91
Edu 4 – Non-Retired	15370	90.09
Edu 5 – Retired	1091	6.39
Edu 5 – Non-Retired	15970	93.61
GS1_5 – Retired	4111	24.10
GS1_5 – Non-Retired	12950	75.90
GS6_9 – Retired	5940	34.82
GS6_9 – Non-Retired	11121	65.18
GS10_13 – Retired	6379	37.39
GS10_13 – Non-Retired	10682	62.61
GS14_15 – Retired	631	3.70
GS14_15 – Non-Retired	16430	96.30
Age 1 – Retired	5868	34.39
Age 1 – Non-Retired	11193	65.61
Age 2 – Retired	8300	48.65
Age 2 – Non-Retired	8761	51.35
Age 3 – Retired	2893	16.96
Age 3 – Non-Retired	14168	83.04
Total – Retired	**2017**	**11.82**
Total – Non-Retired	**15044**	**88.18**

Table F-3
Industrial Trades Discrete Variables

	Frequency	Percentage
Race 1 – Retired	6026	17.43
Race 1 – Non-Retired	28539	82.57
Race 2 – Retired	1563	4.52
Race 2 – Non-Retired	33002	95.48
Race 3 – Retired	302	0.87
Race 3 – Non-Retired	34263	99.13
Race 4 – Retired	4810	13.92
Race 4 – Non-Retired	29755	86.08
Race 5 – Retired	21864	63.25
Race 5 – Non-Retired	12701	36.75
Edu 1 – Retired	3933	11.39
Edu 1 – Non-Retired	30642	88.62
Edu 2 – Retired	18931	54.75
Edu 2 – Non-Retired	15644	45.25
Edu 3 – Retired	11017	31.86
Edu 3 – Non-Retired	23558	68.14
Edu 4 – Retired	597	1.73
Edu 4 – Non-Retired	33978	98.27
Edu 5 – Retired	97	0.28
Edu 5 – Non-Retired	34478	99.72
GS1_5 – Retired	5281	15.27
GS1_5 – Non-Retired	29294	84.73
GS6_9 – Retired	8688	25.13
GS6_9 – Non-Retired	25887	74.87
GS10_13 – Retired	20273	58.63
GS10_13 – Non-Retired	14302	41.37
GS14_15 – Retired	333	0.96
GS14_15 – Non-Retired	34242	99.04
Age 1 – Retired	14482	41.89
Age 1 – Non-Retired	20093	58.11
Age 2 – Retired	14398	41.64
Age 2 – Non-Retired	20177	58.36
Age 3 – Retired	5695	16.47
Age 3 – Non-Retired	28880	83.53
Total – Retired	**3797**	**10.98**
Total – Non-Retired	**30778**	**89.02**

Table F-4
Science & Engineering Discrete Variables

	Frequency	Percentage
Race 1 – Retired	539	3.85
Race 1 – Non-Retired	13461	96.15
Race 2 – Retired	422	3.01
Race 2 – Non-Retired	13578	96.99
Race 3 – Retired	103	0.74
Race 3 – Non-Retired	13897	99.26
Race 4 – Retired	1299	9.28
Race 4 – Non-Retired	12701	90.72
Race 5 – Retired	11637	83.12
Race 5 – Non-Retired	2363	16.88
Edu 1 – Retired	122	0.87
Edu 1 – Non-Retired	13887	99.13
Edu 2 – Retired	3355	23.95
Edu 2 – Non-Retired	10654	76.05
Edu 3 – Retired	4724	33.72
Edu 3 – Non-Retired	9285	66.28
Edu 4 – Retired	3569	25.48
Edu 4 – Non-Retired	10440	74.52
Edu 5 – Retired	2239	15.98
Edu 5 – Non-Retired	11770	84.02
GS1_5 – Retired	1030	7.35
GS1_5 – Non-Retired	12979	92.65
GS6_9 – Retired	1048	7.48
GS6_9 – Non-Retired	12961	92.52
GS10_13 – Retired	11201	79.96
GS10_13 – Non-Retired	2808	20.04
GS14_15 – Retired	730	5.21
GS14_15 – Non-Retired	13279	94.79
Age 1 – Retired	5716	40.80
Age 1 – Non-Retired	8293	59.20
Age 2 – Retired	6393	45.63
Age 2 – Non-Retired	7616	54.37
Age 3 – Retired	1900	13.56
Age 3 – Non-Retired	12109	86.44
Total – Retired	**1557**	**11.11**
Total – Non-Retired	**12452**	**88.89**

Distribution

AIR UNIVERSITY LIBRARY
ARMY MANAGEMENT STAFF COLLEGE LIBRARY
ARMY RESEARCH INSTITUTE LIBRARY
ARMY WAR COLLEGE LIBRARY
CENTER FOR NAVAL ANALYSES LIBRARY
HUMAN RESOURCES DIRECTORATE TECHNICAL LIBRARY
JOINT FORCES STAFF COLLEGE LIBRARY
MARINE CORPS UNIVERSITY LIBRARIES
NATIONAL DEFENSE UNIVERSITY LIBRARY
NAVAL HEALTH RESEARCH CENTER WILKINS BIOMEDICAL LIBRARY
NAVAL POSTGRADUATE SCHOOL DUDLEY KNOX LIBRARY
NAVAL RESEARCH LABORATORY RUTH HOOKER RESEARCH LIBRARY
NAVAL WAR COLLEGE LIBRARY
NAVY PERSONNEL RESEARCH, STUDIES, AND TECHNOLOGY SPISHOCK LIBRARY (3)
PENTAGON LIBRARY
USAF ACADEMY LIBRARY
US COAST GUARD ACADEMY LIBRARY
US MERCHANT MARINE ACADEMY BLAND LIBRARY
US MILITARY ACADEMY AT WEST POINT LIBRARY
US NAVAL ACADEMY NIMITZ LIBRARY